Current Biology 21, 1210–1214, July 26, 2011 ©2011 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2011.06.007

Report

A Chimpanzee Recognizes Synthetic Speech with Significantly Reduced Acoustic Cues to Phonetic Content

Lisa A. Heimbauer,[1,2,*] Michael J. Beran,[1,2] and Michael J. Owren[1,2,*]
[1]Department of Psychology, Georgia State University, PO Box 5010, Atlanta, GA 30302-5010, USA
[2]Language Research Center, Georgia State University, Atlanta, GA 30303, USA

Summary

A long-standing debate concerns whether humans are specialized for speech perception [1–7], which some researchers argue is demonstrated by the ability to understand synthetic speech with significantly reduced acoustic cues to phonetic content [2–4, 7]. We tested a chimpanzee (*Pan troglodytes*) that recognizes 128 spoken words [8, 9], asking whether she could understand such speech. Three experiments presented 48 individual words, with the animal selecting a corresponding visuographic symbol from among four alternatives. Experiment 1 tested spectrally reduced, noise-vocoded (NV) synthesis, originally developed to simulate input received by human cochlear-implant users [10]. Experiment 2 tested "impossibly unspeechlike" [3] sine-wave (SW) synthesis, which reduces speech to just three moving tones [11]. Although receiving only intermittent and noncontingent reward, the chimpanzee performed well above chance level, including when hearing synthetic versions for the first time. Recognition of SW words was least accurate but improved in experiment 3 when natural words in the same session were rewarded. The chimpanzee was more accurate with NV than SW versions, as were 32 human participants hearing these items. The chimpanzee's ability to spontaneously recognize acoustically reduced synthetic words suggests that experience rather than specialization is critical for speech-perception capabilities that some have suggested are uniquely human [12–14].

Results

The current work investigated whether a chimpanzee (*Pan troglodytes*) that comprehends a large set of spoken-English words [8, 9] could also recognize these items when synthesized in forms that significantly reduce acoustic cues to phonetic content present in natural speech [10, 11]. A primary motivation was the issue of whether humans have specialized and potentially unique mechanisms for speech perception. Historically, the argument for specialization grew out of discoveries of evidently speech-specific perceptual phenomena made at Haskins Laboratories in Connecticut, most famously including that human listeners perceive distinctions between stop consonants such as /b/ versus /p/ in an all-or-none, categorical fashion [2]. Resulting claims of a uniquely human speech mode of perception were subsequently challenged, for example when rodents and monkeys trained to discriminate speech sounds were found to show some of the same effects [12, 13]. However, these kinds of experiments have themselves been criticized for requiring prolonged, focused training and involving just a few sounds [4]. Claims of specialization in speech perception remain controversial and unresolved, with prominent voices on both sides of the issue [4–7].

A central problem in trying to understand speech perception in a comparative context is of course that humans are expert language users, whereas animals are not. For example, human experience with phonemes, words, and sentences begins even before birth, critically shaping subsequent perceptual processing and neural organization. Nonhumans do not get such input, making it virtually impossible to know whether a later failure to show a given speech-processing phenomenon reflects the absence of evolved mechanisms or a critical lack of experience with spoken language. Yet, there have been a few, rare cases in which individual animals have received substantial speech experience, which can create unique testing opportunities [15, 16].

The adult chimpanzee Panzee is one such individual, having been reared exclusively by human caregivers from the age of 8 days old. These caregivers treated the chimpanzee much as they would a human infant [17, 18], including routinely speaking to her about functionally relevant topics such as objects, people, and animals in her immediate environment, as well as previous, current, and future activities. Panzee's experience with speech was supplemented by training in a two-way communication system using visuographic, symbol-like lexigrams [19] corresponding to words she was hearing. The current work took advantage of Panzee's resulting abilities to recognize and report spoken-English words by testing her with two unusual forms of synthetic speech (Figure 1 illustrates both natural and synthetic versions; see also Audio S1 and Audio S2 available online).

The first was noise-vocoded (NV) speech, developed as a research tool for simulating the noise-based input received by human cochlear-implant users. NV synthesis retains primarily temporal cueing while severely degrading frequency-specific (spectral) energy distribution [10]. Spectral characteristics such as harmonic structure, formants (vocal-tract resonances), and formant transitions are eliminated or fundamentally altered [10], thereby removing acoustic features otherwise considered to be important acoustic correlates of phonetic content [20, 21]. The second was sine-wave (SW) speech, consisting of just three individual tones that track estimated frequency and amplitude patterns of the lowest formants [11]. Spectral energy is again greatly reduced, but here to three sharp peaks that have been described as an abstracted rendition of natural speech [3, 22]. Recognition of SW speech in particular has been suggested to require specialization, including either an evolved, speech-related brain module [2, 4], or at least a speech mode of perception that is fundamentally different from generalized auditory processing that humans share with nonhumans [3]. To some theorists, differences between natural speech and the spectrally reduced acoustics of NV and SW speech are so extensive as to show that "phonetic perception does not require speech cues" [22].

*Correspondence: lisa.heimbauer@gmail.com (L.A.H.), owren@gsu.edu (M.J.O.)

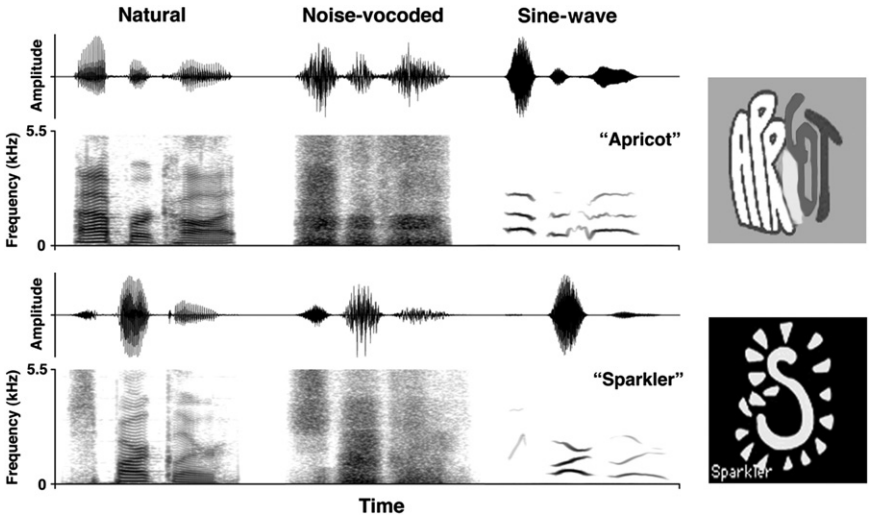

Figure 1. Two Stimulus and Lexigram Examples

Waveforms and narrow-band spectrograms of the words "apricot" and "sparkler" in natural, noise-vocoded (NV), and sine-wave (SW) forms, along with corresponding lexigrams. Waveforms show pressure variation over time, with Fourier transform-based spectrograms revealing corresponding spectral features in the frequency domain (created using a sampling rate of 22.05 kHz and 0.03 s Gaussian analysis window). Both synthetic forms are comprehensible to human listeners but are acoustically significantly reduced relative to natural versions. NV speech retains primarily temporal cues and only rudimentary spectral information—acoustic features such as harmonic structure, formants (vocal-tract resonances), and formant transitions are removed or fundamentally altered [10]. SW speech has just three pure tones that track the lowest formants of natural speech [11], becoming so different that some characterize this form as bearing only an abstract resemblance to the original [3, 22].

The question addressed here was whether Panzee could also understand speech synthesized in these forms. Three experiments tested her with 48 familiar words drawn from an overall recognition vocabulary of 128 items. The words were divided into two sets of 24, one composed of three-, four-, and five-syllable words and one composed of two-syllable items only (stimuli are listed in Table S1, and stimulus balancing issues are discussed in Supplemental Experimental Procedures). Each experiment tested both word sets for three sessions each, with a session consisting of four blocks of 16 natural and 8 synthetic words presented in randomized order. Through systematic rotation across sessions, words were presented a total of 8 times in natural and 4 times in synthetic form (a total of 384 natural and 192 synthetic trials over six sessions). Panzee used a joystick to initiate each trial and then, after hearing the target word twice, to select the corresponding lexigram from among four options shown on a computer monitor ([19]; see Figure 2). In the first two experiments, she was rewarded noncontingently every three to six trials, determined randomly. The experimenter providing the reward remained blind both to lexigram options and to Panzee's choices.

Chimpanzee Outcomes

Results are illustrated in Figure 3, which shows values based on tabulating percentage-correct scores separately by word before computing overall means. Additional analyses ruling out that Panzee was simply cued by syllable number and word duration are included in Supplemental Experimental Procedures. For each experiment, two t tests and a binomial test compared her results against corresponding chance rates of 25% correct, evaluated using Bonferroni-corrected α values of 0.017 [23]. Because below-chance performance could not be plausibly expected, this testing was one-tailed.

The chimpanzee has historically performed at 75% to 85% correct with her familiar spoken words [9]. She was somewhat less accurate here, but comparable and far above chance levels with natural words in both NV [$t(47) = 19.5$, $p < 0.0001$] and SW experiments [$t(47) = 14.8$, $p < 0.0001$]. Panzee's overall percentage correct with NV words was just over 55%, again well above the chance rate, $t(47) = 6.35$, $p < 0.0001$. She was furthermore correct on 26 of 48 (54.2%) first trials with these stimuli, meaning the 48 instances in which she first heard a word in NV form (binomial test, $p < 0.0001$). The chimpanzee was also clearly above chance level with SW stimuli, at about

Figure 2. The Chimpanzee Subject

The chimpanzee subject, Panzee, shown at about 33 months of age (left) and as an adult performing the experimental task (right). She routinely uses a lexigram board for two-way communication with human caregivers and is able to select corresponding lexigrams when hearing 128 different, spoken words.

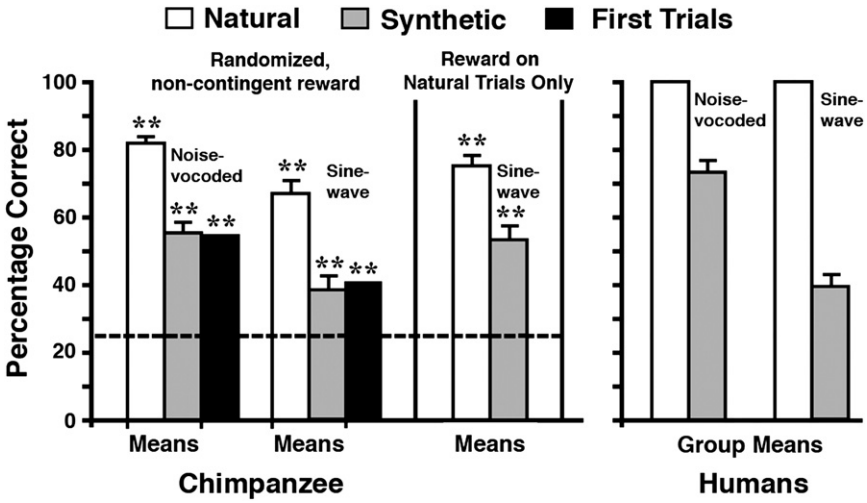

Figure 3. Performance by the Chimpanzee and Human Listeners

Means and standard errors of percentage-correct performance for 48 words heard in natural, NV, and SW forms. Experiments with the chimpanzee, Panzee, included testing each of 48 words 16 times in natural and 4 times in synthetic form. First trials represent the 48 first instances of the chimpanzee hearing a word in a given synthetic form. The first set of SW results shows performance with noncontingent, intermittent reward delivery and no response feedback. The second set shows performance with contingent reward received on natural trials but with no reward or response feedback on SW trials. The dashed line indicates the chance-performance rate of 25% correct. Humans heard and identified all 48 words once each in natural form, followed by either NV (16 listeners) or SW (16 listeners) versions. All comparisons to chance performance were statistically significant at $p \leq 0.008$ and are marked by a pair of asterisks.

40% correct, $t(47) = 2.85$, $p = 0.003$. She appeared frustrated by these sounds but was nonetheless correct on a statistically significant 20 of 48 (41.7%) first trials (binomial test, $p = 0.008$). To verify her performance, we conducted an additional experiment of six sessions with the same sine-wave stimuli 3 months later. This testing used a friendlier regimen providing auditory feedback and immediate reward for correct choices on natural trials, but still no feedback or reward on SW trials. As also shown in Figure 3, Panzee's performance with the SW stimuli then rose to about 53% correct [$t(47) = 6.24$, $p < 0.0001$], likely reflecting increased motivation to perform the task.

Finally, a nested, one-way analysis of variance (ANOVA) with Bonferroni correction and log transformation revealed no overall performance differences across the three experiments but showed an interaction between experiment and synthesis factors, $F(3,282) = 20.4$, $p < 0.001$. Two-tailed, planned pairwise comparisons revealed equivalent performance with natural stimuli across NV and both SW experiments, while confirming lower initial performance on SW than NV versions ($p = 0.03$), as well as better performance in the second than in the first experiment with SW versions ($p = 0.002$).

Human Outcomes

Thirty-two human participants were tested for comparison, hearing and transcribing the 48 words one at a time first in natural and then in NV or SW form. As shown in Figure 3, mean performance exceeded 99% correct for natural words in both groups. However, performance with synthetic items paralleled Panzee's results. One-way nested ANOVA with Bonferroni correction revealed no overall difference between the two experiments with humans, but a significant experiment-by-synthesis interaction, $F(2,188) = 55.2$, $p < 0.0001$. Although a substantial proportion of the words were correctly transcribed in both synthesis conditions, two-tailed, planned pairwise comparisons showed that NV stimuli were less accurately identified than natural words, with lower performance with SW stimuli ($p < 0.03$ in both cases).

Discussion

Overall, we conclude that the chimpanzee subject demonstrated functional perception of both NV and SW speech. Although two previous studies have tested NV and SW phoneme discrimination in chinchillas [24] and birds [25], respectively, Panzee is the first nonhuman demonstrated to show word-level recognition with either synthesis method. A notable feature of the testing was that this animal was willing to work while receiving only intermittent, noncontingent food reward. The experimenter could therefore remain blind both to the lexigram options and to Panzee's choices on critical test trials, ruling out any possibility that the experimenter was providing inadvertent cueing or that the animal was learning the correct responses for synthetic stimuli.

Performance with Synthetic Words

Although she had never previously heard words in NV form, Panzee's performance with these items was convincing from the initial first trials on. She was less accurate than with natural versions, but that was also the case for the human listeners. Human performance was comparable to that found in other work, which has furthermore shown that isolated NV words are more difficult to understand than sentence-length stimuli [26, 27]. Panzee's performance when initially hearing SW items was also significantly above chance level, although both she and the humans scored lower than with NV words. SW speech is challenging, and not every human listener readily recognizes it. Studies presenting sine-wave sentences have produced variable outcomes, for instance ranging from about 30% [28] to about 85% correct [3, 22]. In one other study that tested participants transcribing individual sine-wave words, the mean outcome was about 52% correct [29]. A further test of the same words using a four-alternative forced-choice procedure (printed words; 25% chance rate) produced about 60% correct performance. Participants in the current experiments received minimal practice with SW speech, while falling well within the range of variation of previous studies at about 40% correct.

Procedural differences make it difficult to directly compare performances by Panzee and the human listeners. However, it is clear that both parties were more challenged by SW than NV items, while identifying significant proportions of the words in each case. The chimpanzee's recognition of SW words was confirmed by the second test with these stimuli, conducted months later with no intervening exposure. There she was rewarded more frequently, although on natural trials only. Her improvement to over 50% correct with SW items could

therefore not have resulted from reinforcement of particular responses, instead likely reflecting increased effort, greater familiarity with SW stimuli, or both.

Implications for Specialized Perception of Speech

Human speech perception is a multifaceted process in which listeners can capitalize on any of a potential variety of cues in a given circumstance. Whether this kind of flexible, robust perception implies specialized processing is subject to debate, and the significance of NV and SW speech for that discussion is by no means a matter of universal agreement. We nonetheless suggest that the current work is informative in demonstrating that a chimpanzee was also able to identify many familiar words presented in these forms, doing so upon first exposure, without explicit training, and without differential feedback. Her relative performance levels for natural, NV, and SW words were also paralleled by human participants.

On the one hand, these findings do not show equivalence of processing in humans and the chimpanzee, nor do they rule out the possibility that humans are specialized for speech perception. On the other hand, Panzee's performance does suggest that even quite sophisticated human speech perception phenomena may be within reach for some nonhumans. Such outcomes might, for example, reflect vocal-perception mechanisms shared with chimpanzees or other great apes [30]. Alternatively, effective perception of NV and SW speech may be understandable as a particularly well-developed form of top-down processing—the general cognitive strategy of using preexisting knowledge to interpret inherently ambiguous sensory input [31, 32]. In that case, it is simply not known whether the level of complexity required is possible for many or just a few nonhuman species.

Overall, our findings at least suggest that the perceptual capacities required for speech processing in the reduced acoustic forms tested here may be traceable as far back as the common ancestor of chimpanzees and humans—and perhaps even further. These results also underscore the critical role that experience with spoken language likely plays in human speech perception capabilities, as well as the importance of testing humans and nonhumans on a more level playing field. Finally, although not ruling out that specializations for speech perception may have arisen specifically among ancestral humans, this evidence suggests that any such changes more likely acted to improve upon, rather than to invent, the processing capabilities involved.

Experimental Procedures

Word Selection and Acoustical Methods

Test items were drawn from a set of 128 words on which Panzee had scored 75% or more correct in annual, word-to-lexigram testing conducted the year the experiments began. The first 24-item set presented consisted of three- to five-syllable words, whereas the second included only two-syllable words (see also Supplemental Experimental Procedures). All words were originally spoken by author M.J.B., who was recorded using a Shure PG14/PG30-K7 head-worn wireless microphone, a Realistic 32-12008 stereo mixing console, and a Marantz PMD671 solid-state audio recorder. Acoustical editing and synthesis were performed using Praat freeware (www.praat.org) and custom-written scripts (http://sites.google.com/site/psyvoso). Files were sampled at 22.05 kHz, filtered to remove any 60 Hz energy and DC offset, and normalized to the full 16-bit amplitude range available. Individual words were extracted from the original recordings with onsets and offsets set at zero crossings, and 100 ms silence was then added to the beginning and end of each file.

NV and SW algorithms used were developed by C. Darwin (www.lifesci.sussex.ac.uk/home/Chris_Darwin). NV words were created by first extracting seven frequency-limited energy bands from the original waveform, together spanning a range of 50 to 11025 Hz (i.e., 50–800, 800–1500, 1500–2500, 2500–4000, 4000–6000, 6000–8500, and 8500–11025). Amplitude envelopes were extracted from each of the resulting waveforms and used to temporally modulate corresponding frequency-limited, white-noise bands. These bands were then summed, with leading and trailing silence added to create the final stimuli. SW versions were made by first extracting the frequencies and amplitudes of the three lowest formants across each word. Resulting contours were hand edited as needed to produce the best possible match to formant tracks visible in spectrographic representations of each word and then used to synthesize three corresponding sine waves. These sine waves were summed and amplitude normalized, with leading and trailing silence added to create the final stimuli.

Chimpanzee Testing

Panzee was 22 years old when testing began and housed at the Language Research Center at Georgia State University (GSU). She has a long history of responding to spoken English in informal interactions and formal testing [8, 9, 17–19] but was naive to NV and SW synthesis. Test sessions lasted about 30 min, with Panzee working for rewards such as cherries, blackberries, blueberries, grapes, or other foods. Sessions with natural words were conducted as needed to ensure stable performance before NV and SW testing began. Stimuli were computer-presented using ADS L200 speakers and a 19-inch monitor on a cart outside Panzee's cage (see Figure 2). The test program was written in Visual Basic version 6.0, and the manipulandum was a customized joystick (Gravis 42111 Gamepad Pro) that the chimpanzee used to initiate trials, hearing each stimulus twice. After 1 s, the correct lexigram and three randomly selected foils from the set appeared in randomized screen locations. NV sessions occurred without any prior exposure. Before testing with SW words, eight nontest items from Panzee's larger word set were presented over five informal sessions. These sessions introduced the chimpanzee to interpreting SW sounds as speech, as is also required with humans [11]. Animal maintenance and testing procedures were approved by GSU's Institutional Animal Care and Use Committee and met all national and institutional standards and policies.

Human Testing

Human participants were 32 (8 male, 24 female) university students (age 18 to 39) working for course credit and tested in groups of one to five in a sound-deadened room using TDT System II modules. They heard stimuli presented in randomized order through Sennheiser HD650 headphones and wrote down their responses within 8 s. All participants were introduced to either NV or SW speech when their respective sessions began, hearing a synthesized male voice counting from one to ten and ten to one. This sequence was repeated until all participants present reported perceiving these sounds as speech. Transcribed items were scored as being correct if and only if the response exactly matched the word presented, although homonyms and obvious misspellings of the word were considered correct. All participants provided informed consent prior to testing. Procedures were approved by GSU's Institutional Review Board and met all national and institutional standards and policies.

Supplemental Information

Supplemental Information includes one table, Supplemental Experimental Procedures, and two audio files and can be found with this article online at doi:10.1016/j.cub.2011.06.007.

Acknowledgments

We thank M. Hart, J. Kelley, S. Hunsberger, and D. Rice for expert animal care and H. Rakusin, K. Wheeler, and K. Colmer for help in testing human listeners. R. Sevcik, M. Christiansen, J. Sinnott, and three anonymous reviewers provided valuable feedback. Work was supported by the National Institute of Child Health and Human Development, National Science Foundation, GSU's RCALL and Brains & Behavior programs, the Center for Behavioral Neuroscience under the Science and Technology Centers Program of the National Science Foundation under agreement IBN-9876754, and ERC grant agreement AdG 249516. L.A.H. was funded by an RCALL Fellowship and a Duane M. Rumbaugh Fellowship.

Received: November 29, 2010
Revised: May 3, 2011
Accepted: June 3, 2011
Published online: June 30, 2011

References

1. Liberman, A.M., Cooper, F.S., Shankweiler, D.P., and Studdert-Kennedy, M. (1967). Perception of the speech code. Psychol. Rev. *74*, 431–461.
2. Liberman, A.M., and Mattingly, I.G. (1985). The motor theory of speech perception revised. Cognition *21*, 1–36.
3. Remez, R.E., Rubin, P.E., Berns, S.M., Pardo, J.S., and Lang, J.M. (1994). On the perceptual organization of speech. Psychol. Rev. *101*, 129–156.
4. Trout, J.D. (2001). The biological basis of speech: What to infer from talking to the animals. Psychol. Rev. *108*, 523–549.
5. Diehl, R.L., Lotto, A.J., and Holt, L.L. (2004). Speech perception. Annu. Rev. Psychol. *55*, 149–179.
6. Fitch, W.T., Hauser, M.D., and Chomsky, N. (2005). The evolution of the language faculty: Clarifications and implications. Cognition *97*, 179–210, discussion 211–225.
7. Pinker, S., and Jackendoff, R. (2005). The faculty of language: What's special about it? Cognition *95*, 201–236.
8. Williams, S.L., Brakke, K., and Savage Rumbaugh, E.S. (1997). Comprehension skills of language-competent and nonlanguage-competent chimpanzees. Lang. Commun. *17*, 301–317.
9. Beran, M.J., Savage-Rumbaugh, E.S., Brakke, K.E., Kelley, J.W., and Rumbaugh, D.M. (1998). Symbol comprehension and learning: A "vocabulary" test of three chimpanzees. Evol. Commun. *2*, 171–188.
10. Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. Science *270*, 303–304.
11. Remez, R.E., Rubin, P.E., Pisoni, D.B., and Carrell, T.D. (1981). Speech perception without traditional speech cues. Science *212*, 947–949.
12. Brown, C.H., and Sinnott, J.M. (2006). Cross-species comparisons of vocal perception. In Listening to Speech: An Auditory Perspective, S. Greenberg and W.A. Ainsworth, eds. (Mahwah, NJ: Erlbaum Associates), pp. 183–202.
13. Kluender, K.R., Lotto, A.J., and Holt, L.L. (2006). Contributions of nonhuman animal models to understanding human speech perception. In Listening to Speech: An Auditory Perspective, S. Greenberg and W.A. Ainsworth, eds. (Mahwah, NJ: Erlbaum Associates), pp. 203–220.
14. Holt, L.L., and Lotto, A.J. (2008). Speech perception within an auditory cognitive science framework. Curr. Dir. Psychol. Sci. *17*, 42–46.
15. Pepperberg, I.M. (2010). Vocal learning in Grey parrots: A brief review of perception, production, and cross-species comparisons. Brain Lang. *115*, 81–91.
16. Rumbaugh, D.M., Beran, M.J., and Savage-Rumbaugh, E.S. (2003). Language. In Primate Psychology, D. Maestripieri, ed. (Cambridge, MA: Harvard University Press), pp. 395–423.
17. Brakke, K.E., and Savage-Rumbaugh, E.S. (1995). The development of language skills in bonobo and chimpanzee—I. Comprehension. Lang. Commun. *15*, 121–148.
18. Brakke, K.E., and Savage-Rumbaugh, E.S. (1996). The development of language skills in *Pan*—II. Production. Lang. Commun. *16*, 361–380.
19. Rumbaugh, D.M., and Washburn, D.A. (2003). Intelligence of Apes and Other Rational Beings (New Haven, CT: Yale University Press).
20. Raphael, L.J. (2005). Acoustic cue to the perception of segmental phonemes. In Handbook of Speech Perception, D.B. Pisoni and R.E. Remez, eds. (Oxford: Blackwell Scientific), pp. 182–206.
21. Stevens, K.N. (2000). Acoustic Phonetics (Cambridge, MA: MIT Press).
22. Remez, R.E. (2005). The perceptual organization of speech. In Handbook of Speech Perception, D.B. Pisoni and R.E. Remez, eds. (Oxford: Blackwell Scientific), pp. 9–26.
23. Bakeman, R., and Robinson, B. (2005). Understanding Statistics in the Behavioral Sciences (Mahwah, NJ: Erlbaum Associates).
24. Loebach, J.L., and Wickesberg, R.E. (2006). The representation of noise vocoded speech in the auditory nerve of the chinchilla: Physiological correlates of the perception of spectrally reduced speech. Hear. Res. *213*, 130–144.
25. Dooling, R.J., Best, C.T., and Brown, S.D. (1995). Discrimination of synthetic full-formant and sinewave /ra-la/ continua by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taeniopygia guttata*). J. Acoust. Soc. Am. *97*, 1839–1846.
26. Hervais-Adelman, A., Davis, M.H., Johnsrude, I.S., and Carlyon, R.P. (2008). Perceptual learning of noise vocoded words: Effects of feedback and lexicality. J. Exp. Psychol. Hum. Percept. Perform. *34*, 460–474.
27. Loebach, J.L., and Pisoni, D.B. (2008). Perceptual learning of spectrally degraded speech and environmental sounds. J. Acoust. Soc. Am. *123*, 1126–1139.
28. Saldaña, H.M., Pisoni, D.B., Fellowes, J.M., and Remez, R.E. (1996). Audio-visual speech perception without speech cues. In Fourth International Conference on Spoken Language Proceedings. *4*, pp. 2187–2190.
29. Liebenthal, E., Binder, J.R., Piorkowski, R.L., and Remez, R.E. (2003). Short-term reorganization of auditory analysis induced by phonetic experience. J. Cogn. Neurosci. *15*, 549–558.
30. Taglialatela, J.P., Russell, J.L., Schaeffer, J.A., and Hopkins, W.D. (2008). Communicative signaling activates 'Broca's' homolog in chimpanzees. Curr. Biol. *18*, 343–348.
31. Plomp, R. (2001). The Intelligent Ear (Mahwah, NJ: Erlbaum Associates).
32. Davis, M.H., and Johnsrude, I.S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. Hear. Res. *229*, 132–147.